INTELLECTUAL

Espacenet

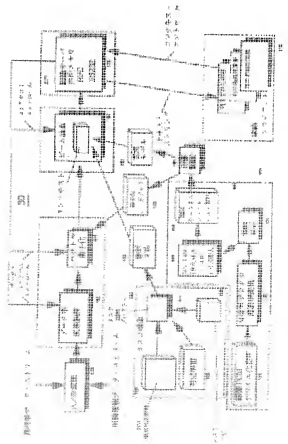
Bibliographic data: JP6204952 (A) — 1994-07-22

TRAINING OF SPEECH RECOGNITION SYSTEM UTILIZING TELEPHONE LINE

Inventor(s):		BINSU EMU SUTANFUOODO; NOOMAN EFU BURITSUKUMAN ±
Applicant(s):		IBM ±
Classification:	**- international:**	***G06F3/16; G10L15/00; G10L15/06; H04B14/04; G10L15/02; G10L15/14; G10L15/20; G10L19/00;*** (IPC1-7): G06F3/16; G10L5/06; H04B14/04
	- european:	G10L15/06T
Application number:		JP19930219208 19930812
Priority number(s):		US19920948031 19920921
Also published as:		JP2524472 (B2) US5475792 (A)

Abstract of JP6204952 (A)

PURPOSE: To provide a system with which continuous voices inputted from a telephone line can be recognized by inputting a correction voice data set to a voice recognizing processor and training a statistical pattern matching unit. CONSTITUTION: A digital voice input from an A/D 100 is sampled by a data rate converting device 102, a voice data stream is divided into fine frames by a vector-quantizing block 104, and characteristics are extracted from the respective frames and summerized. Vector quantization is similarly used in a training process 190, as well. A quantizing code book 105 preserves a code book for telephone voice generated by the function 190. On the other hand, input voice data 180 are resampled in a block 182, supplied to a code/decode band filter and set to a prescribed band width. The correction voice data set is inputted to the voice-recognizing processor, and the statistical pattern matching unit is trained. The voice recognition is executed, while using a voice recognition processor from a telephone system.



Last updated 5.12.2011 Worldwide Database 5.7.31; 93p

(19)日本国特許庁（JP）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平6－204952

(43)公開日　平成6年(1994)7月22日

(51)Int.Cl.[5]	識別記号	庁内整理番号	FI	技術表示箇所
H04B 14/04		Z 4101－5K		
G06F 3/16	310	A 7165－5B		
G10L 5/06		Z 9379－5H		

審査請求　有　　請求項の数7（全 13 頁）

(21)出願番号　特願平5－219208

(22)出願日　平成5年(1993)8月12日

(31)優先権主張番号　07／948，031

(32)優先日　1992年9月21日

(33)優先権主張国　米国（US）

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレイション
INTERNATIONAL BUSINESS MASCHINES CORPORATION
アメリカ合衆国10504、ニューヨーク州　アーモンク　（番地なし）

(72)発明者　ビンス　エム　スタンフォード
アメリカ合衆国メリーランド州ゲイサスバーグ　ソラコンブリッジ　テラス　15349番地

(74)代理人　弁理士　合田　潔　（外2名）

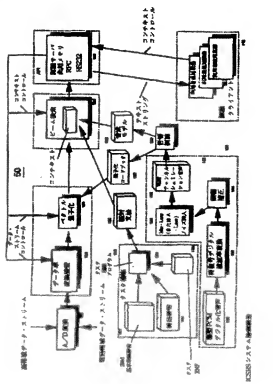
最終頁に続く

(54)【発明の名称】　電話回線利用の音声認識システムを訓練する方法

(57)【要約】

【目的】　本発明は、電話回線から入力される連続音声を音声認識できるシステムを提供する。

【構成】　本発明は、電話システムからの音声に対応できる音声認識システムを訓練するための電話チャネル・シミュレーションを開示する。電話音声帯域より高い帯域を持つ音声認識訓練プロセッサに音声データ・セットが入力される。入力音声データ・セットを修正する一連の処理が行われ、最終的に修正される音声データ・セットによって音声認識プロセッサが電話からの音声信号の音声認識を行うことが可能となる。



1

【特許請求の範囲】

【請求項1】　電話帯域幅より広い帯域幅の音声認識訓練プロセッサへの音声データ・セットの入力ステップと、

上記音声データ・セットを間引き、上記電話帯域幅を有する間引かれた音声データ・セットを入手するステップと、

帯域通過デジタル沪波器を上記間引かれた音声データ・セットに適用し、電話機器の電送特性に特性化した、沪波された音声データ・セットを入手するステップと、

上記沪波された音声データ・セットの振幅を、その最大ダイナミック・レンジが非圧伸電話音声の最大ダイナミック・レンジと一致するように補正し、振幅補正音声データ・セットを入手するステップと、

上記振幅補正音声データ・セットを、電話システムの圧伸・非圧伸音声信号シークエンスを表す量子化ノイズを用いて修正し、修正音声データ・セットを入手するステップと、

上記修正音声データ・セットを音声認識プロセッサに入力し、統計的パターン・マッチング・ユニットを訓練するステップと、

から構成される、電話システムから得られる音声に応答する音声認識プロセッサを訓練する方法。

【請求項2】　上記電話帯域幅が上記音声幅の高位帯域より低い帯域である上記請求項1記載の方法。

【請求項3】　上記帯域通過デジタル沪波器が最大平坦設計アルゴリズムを備え持つ上記請求項1記載の方法。

【請求項4】　上記音声データ・セット振幅補正の結果、最大ダイナミック・レンジが非圧伸ｍｕ－ｌａｗ電話音声の最大ダイナミック・レンジに一致する上記請求項1の記載方法。

【請求項5】　上記音声データ・セット振幅補正の結果、最大ダイナミック・レンジが非圧伸Ａ－ｌａｗ電話音声の最大ダイナミック・レンジに一致する上記請求項1記載の方法。

【請求項6】　上記音声データ・セット修正ステップがｍｕ－ｌａｗノイズとしての量子化ノイズを用いる上記請求項1記載の方法。

【請求項7】　上記音声データ・セット修正ステップがＡ－ｌａｗノイズとしての量子化ノイズを用いる上記請求項1記載の方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、公衆電話交換回線を利用する音声認識システムに関するものである。

【０００２】

【従来の技術】音声認識システムは、よく知られている技術である。ＩＢＭタンゴラ(Tangora)[13]（本願書文末記載の参照文献の番号で、以下同様に表記する）およびドラゴン・システム・ドラゴン３０ｋ以上述システムはその例である。それらは、典型的な単一ユーザおよび話し手依存型システムである。これは、プロセスが「登録」と呼ばれるプロセスの間に、話し手の音声パターンで音声認識装置を訓練することを各話し手に要求する。将来の認識セッションの中で話し手自身をシステムが識別しなければならないのでシステムは話し手のプロファイルを維持する。典型的には、話し手は低レベル雑音システム環境の中でローカル・マイクを通して認識システムが常駐する単一の機械に話しかけながら登録を行う。登録作業の間、その話し手は、長ったらしい原稿を読むことを要求されるが、それ故に、そのシステムは各話し手の特色に順応することができることとなる。　独立した口述システム（たとえば、上記の２つのシステム）は、話し手にたどたどしい、不自然な形で、すなわち、語と語の間にポーズをいれながら、各語を形づくることを要求する。これにより、音声認識システムは、語の境界となる、先行および後続の無音を利用し、各個人の話に連想される音声パターンを識別することが可能となる。典型的音声認識システムは、（たとえば、ＩＢＭタンゴラ システムの Office Correspondence の場合のように）単一の機械上で作動し、訓練された単一の適用業務を持つ。

【０００３】話し手依存型音声認識装置をもつマルチ・ユーザ・システム環境は、各話し手にその音声パターンをシステムに理解させるための退屈な訓練に従事することを要求する。話し手の電話番号によってシステムがどの音声テンプレートを使用すべきかを知り得る共通データ・ベースに音声テンプレートが格納されているかもしれないが、それでもなお各話し手は使用の前にそのシステムを訓練しなければならない。外部の電話線から接続してくる新しいシステム利用者は、このプロシージャが容認できるものでないことを認識する。また、成功した電話の音声認識システムというものは、様々の分野に関係する音声を正確に認識するために迅速な文脈切り替えができなければならない。たとえば、一般のオフィス通信のために訓練されたシステムは、数字列の提示の場合、うまく働かない。

【０００４】Kai-Fu Lee の博士号論文[1]の中で最初に記述されたスフィンクス(Sphinx)システムは、以前の話し手依存型認識システムに大きな進歩をもたらした。それは話し手独立型であり、会話音声の連続ストリームから単語を認識することができた。このシステムは、実際の使用に先立って行われる話し手個々の登録を必要としなかった。話し手依存型システムの中には、話し手に４～６週毎に再登録することを要求したり、利用者にそのシステムが理解するための個人用プラグイン・カートリッジを持ち運ぶことを要求する、連続の音声認識を行うスフィンクス・システムは、語と語の間の休止を必要とせず、音声認識システムの一時的ユーザに非常に多くの親切なアプローチを提供する。認識システムの利便のた

2

3

めにどのように音声を調節するかをユーザが訓練しなくてよいので、この点は、電話の音声認識システムの本質的な特長である。

【0005】音声認識システムは、また、与えられたささやかな語彙を使って、実時間処理を提供しなければならない。しかし、スフィンクス・システムは、まだ以前の話し手に依存する認識システムの不利な点をいくつか持っていた。マイクロホンおよび比較的に制約された語彙を使用しながら低レベル雑音システム環境の中で単一機械上で操作するようプログラムされていた。スフィンクス・システムは、複数ユーザのサポート、少くとも、異なるロケーションおよび複数の語彙認識に関するサポートを行うようには設計されなかった。

【0006】

【発明が解決しようとしている課題】本発明は、上記の従前技術の不利益な点の多くを克服することを目的とする。したがって、本発明は、ローカルおよび遠隔地双方の話し手からの入力を持つ電話機器使用に適した連続音声話し手独立型音声認識装システムを提供することを目的とする。

【0007】低レベル雑音条件の中で集められた語彙を基に電話システム環境のように高レベル雑音の中での音声パターンを認識できるようシステムを訓練することは、本発明のもうひとつの目的である。

【0008】複数の音声適用業務が、コンピュータ・ネットワーク上または電話線上で同時に音声認識システムによって音声認識されるようにすることは、本発明のもうひとつの目的である。

【0009】

【課題を解決するための手段】本発明の上記目的は、ローカル・エリア・ネットワークまたは広域ネットワークの上のクライアント・サーバを基に構築される音声認識システムによって達成される。この音声認識システムは、アナログまたはディジタル音声データを音声を表わす一組のケプストラム係数およびベクトル量子化値に変換するフロントエンドを含む多くのモジュールに分けられる。バックエンドは、ベクトル量子化値を使用して、その音声の作る文脈と音素モデル(Phoneme Models)と語対文法(Word Pair Grammars)に従ってその語を認識する。語彙を一連の文脈（すなわち、ある特定の語がそのシステムによって予期される状況）に分割することによって、一層大規模な語彙を、最小限のメモリに収納することができる。ユーザが音声認識作業を進めるにつれて、文脈は共通のデータベースから迅速に切り換えられる(下記引用Brickmanその他による特許出願参照)。システムは、また、コンピュータ・ネットワーク間および複数のユーザ適用業務間のインターフェースを備える。

【0010】このシステムは、文脈のための語対文法を構築しシステムを訓練する訓練およびタスク構築モジュールを備え持つ。

4

【0011】本発明は、電話から得られる音声に応答できるよう音声認識システムを訓練するための電話チャネル・シミュレーション・プロセスを含む。その方法は、音声データ・セットを、電話帯域幅より帯域幅が高い音声認識訓練・プロセッサに入力することから始める。入力音声データ・セットは、電話帯域幅を持つ間引かれた(decimated)音声データを得るために間引かれる。その後、帯域通過デジタル沢波器(Bandpass Digital Filter)を間引かれた音声データ・セットに適用し、電話機器の伝送特性に特化させる。これにより、沢波された音声データ・セットが得られる。次に、その沢波された音声データ・セットの最大ダイナミック・レンジが非圧伸(uncompanded)電話音声の最大レンジと一致するように、振幅補正（スケーリング）を行う。それから、振幅補正された音声データ・セットは、電話システムの音声信号を圧伸、非圧伸するシークエンスを表わす量子化ノイズによって修正される。そして修正された音声データは、統計的パターン・マッチング・データ・ユニットを訓練するために音声認識プロセッサに入力される。上記方法により、音声認識プロセッサは電話システムからの音声信号に対して音声認識を実行することができることとなる。

【0012】

【実施例】電話ラインによってもらされる帯域幅減衰および雑音は、すべての音声認識システムの正確度を減ずる。この影響は、瞬間的に認識されなければならない語彙の大きさに応じて増加する。迅速に切り替え可能な音声認識文脈の使用はこの発明にとって有用であるが、そのために、個々の文脈のサイズが制限されねばならない。文脈切り替えは、N.F. Brickmanその他の発明になるIBM出願のアメリカ合衆国特許出願番号947,654"Instantaneous Context Switching For Speech Recognition Systems"で記述されるており、本明細書においても参照される。図1は、ハードウェア機械構成から独立したIBM連続音声認識システム（IBM Continuous Speech Recognition System のことで、以下ICSRSと呼ぶ）の論理的構造を図示する。ICSRSは、幅広いレベルで、以下の分野をカバーするコンポーネントから構成される。

【0013】データ収集：データは、アナログからディジタル形式にブロック100で変換されるか、あるいは電話のデータの場合他のチャネルから潜在的にデマルチプレックス(demultiplexed)される。

【0014】データ圧縮：ICSRSフロントエンドブロック102および104は、ベクトル量子化ステップの間に300バイト/秒に音声データストリームを調整し、再標本化し、圧縮する。

【0015】音声認識：バックエンド106は、文法ガイド型ビーム・サーチ・アルゴリズムを使用しているパターンマッチング音素モデル192によって実際の音声

5

認識を実行する。音素モデル１９２および語対文法１３５は共に認識文脈を構成する。バックエンド認識装置のひとつまたは複数の事例が、遠隔地であろうがローカルであろうが音声データを捕捉し圧縮するフロントエンド事例に配備されることができる。

【００１６】タスク構築：タスク構築コンポーネント１３０は、認識文脈のオフラインでの構築を可能にし、実行時で使用のために語対文法をコンパイルし、適切な音素モデルをそのタスク（文脈）に連結させる。

【００１７】適用業務プログラム・インタフェース（ＡＰＩ）：ＡＰＩ１０８は、データストリーム・コントロール、文脈ローディングおよび起動を可能にするＲＰＣ(Remote Procedure Call)に基づく認識サービスを提供する。

【００１８】電話チャネル・シミュレータ：シミュレータ１８５は、高帯域、高解像度音声データ・セットを、音素モデル１９２および電話音声に連結し、 減少された標本抽出率、圧縮された帯域幅および圧縮されたダイナミック・レンジの電話音声を作り出す。

【００１９】音声認識の間に、ローカル・マイクからの高帯域声データストリームも電話に関連しているような低帯域声データストリームも、アナログデジタル変換ブロック１００によって受け取られる。アナログデジタル変換１００は、ボイス・ワークステーション上のIBM M-Audio Capture/Playback Cardカード（Ｍ－ＡＣＰＡ）のようなハードウェア・カードによって実行されることができる。Ｍ－ＡＣＰＡは、高帯域または電話帯域幅信号を処理するデジタル信号処理機構を持ち、ディジタルに標本化された一連のデータ・ポイントにそれらを変換する。この変換は、また、デジタルＰＢＸや８ＫＨｚ、８ビットのＭｕ－Ｌａｗ／Ａ－Ｌａｗ形式で与えられる電話データストリームによって実行されることもできる。

【００２０】本発明では、高帯域を、サンプル率１６キロヘルツ以上と定義する。低レベル帯域幅を、アメリカ合衆国で一般の電話がデジタル音声に使う８キロヘルツ以下と定義する。電話システムの中でデジタル情報が個人の電話交換（ＰＢＸ）から入る可能性があるので、Ａ／Ｄ変換ブロック１００は、オプションとして必要である。

【００２１】音声認識に対する「フロントエンド」の中の最初の重要なブロックは、データ条件付け・速度変換ブロック１０２(Data Conditioning and Rate Conversionのことで、以下ＤＣＲＣと呼ぶ)である。Ａ／Ｄ変換１００からのデジタル化された入力は、４４または８ＫＨｚである。本発明で間引き(DECIMATION)と呼び使用する再標本化テクニックは、ＩＥＥＥ の文献[2]によって提供されている。ＤＣＲＣ１０２は、デジタル化された信号に対しアンチエイリアシング(Anti-aliasing)・フィルターを使用し標本化を行い、次のステップでの使用のために、１６ＫＨｚまたは８ＫＨｚデータストリームを作る。ＤＣＲＣおよびベクトル量子化プロセスは、以下に詳細に記述される。

【００２２】音声認識の中でデータ条件付け・速度変換の後、音声データは、ベクトル量子化ブロック１０４に渡される。ベクトル量子化の中でデジタル・データ・ストリームは、１秒間の１／５０のフレームに細分化され、１６ＫＨｚ、１１ＫＨｚおよび８ＫＨｚそれぞれの標本化率に対し各々３２０個、２２０個および１６０個の標本となる。本発明の好ましい実施例のひとつでは、いかなる帯域幅音声信号からも計算される１秒につき１００フレームがあり、それらは５０パーセント上重ねされ、ハミング・ウィンドウ(Hamming Window)が適用される。ハミング・ウィンドウは、文献[3]で定義されている。

【００２３】音声データストリームがフレームに細分化されたあと、ベクトル量子化ステップは、各フレームから特性を抽出する。ベクトル量子化ステップの抽出部分で、ＬＰＣケプストラム係数と呼ばれる一連のパラメータが、計算される。ケプストラム係数は、パターン認識のために音声の重要な特性のいくつかを抜き出し、要約する。データの各フレームの中で、音声の１秒の５０分の１が、カプセルに入れられる。１秒につき５０のフレームと想定するであろうが、５０パーセントの上重ねがあるので、１秒につき１００フレームが生成される。ケプストラム係数を計算するために、まず（コサイン・ベル-cosine bell-である）ハミング・ウィンドウが、音声データに適用される。抽出されたデータが、無限時間連続フーリエ変換にあるようにするために、ハミング・ウィンドウは、音声データの各フレームのエッジを次第に減少させる。

【００２４】ハミング・ウィンドウ化されたフレームは音声スペクトルを平坦にするために、そのＺ変換が１．０－０．９７＊z^{-1}（[1]49ページ参照）であるところの沪波器を使用して事前に沪波される。それから、１４個の自己相関係数が計算される。自己相関係数が、文献[4]の記述でよく知られている方法でケプストラム係数を計算するために使われる。１３個のケプストラム係数は、１４個の自己相関係数から引き出される。自己相関係数の数やケプストラム係数の次元数を変えることは可能である。これらの係数の統計的特性は、最終的なベクトル量子化ステップをガイドするために使われる。

【００２５】ベクトル量子化は、訓練プロセス１９０の中でも同様に使われる。下記の訓練データの調整は、基本スフィンクス認識エンジンを電話機器上で作動可能とさせる点で、本発明にとって重要である。訓練プロセス１９０において、１０，０００から１５，０００の間のセンテンスがとられて、フレームに細分化され、そこから自己相関およびケプストラム係数が計算される。参照文献[5]に記述されるｋ－手法タイプのクラスタリング

・プロシージャを使用して、256個のクラスにケプストラム・フレーム特性を区分する。これらのケプストラム・クラスターの中央値、およびそのクラス・ラベルが共に取り出され、これ以後「コード・ブック」と呼ばれる。量子化コード・ブック105は、音響訓練機能190によって生成される電話音声用コード・ブックを保存し、また、第2の高帯域音声用コード・ブックをも保存する。

【0026】ベクトル量子化の最終的なステップのために、どのクラスター中央値がフレーム・ケプストラム係数に最も近いかを決めるために、ブロック104は、上記のように訓練プロシージャで引き出される量子化コード・ブック105のコード・ブックを参照する。現在のフレームが、コード・ブック値によって表わされたクラスに割り当てられる。256個のクラスがあるので、VQ(Vector Quantization)値は、1バイトで表わされる。微分ケプストラムおよびフレームのそのべき乗から引き出される別の2個の1バイトVQ値がある。1秒に100回引き出されるこの3個の1バイトVQ値があり、その結果、音声データストリームは2、400 ビット／秒に圧縮される。

【0027】音声認識装置のためにその音声を特徴づけるところの完全に別個のコード・ブックが、電話データから引き出され、図1の量子化コード・ブック105で保存されなければならないということは、電話音声認識に関する本発明の一部である。また、対応する音素モデルが電話データから引き出され、音素モデル192で保存されなければならないということは、本発明のもう一つの部分である。標本率減少、帯域幅圧縮およびダイナミック・レンジ圧縮のために、電話音声信号はかなり変わる。しかし、多大な労力を要する、電話から収集する音声標本の使用を必要とせず、高帯域標本を、電話チャネル特性をシミュレートするように処理することができる。これにより、スフィンクス・システムの初期化訓練で使われた、大規模で既に使用可能の音声データ・ファイルを活用して、電話音声認識を可能とする。電話チャネル・シミュレータは、本発明の対象である。

【0028】電話チャネル・シミュレーションは、下記の通り、3つの段階的プロセスで達成される。

1.) 電話帯域幅への変換

文献[14]から[19]で参照されるように、(たとえば44、100Hz、あるいは16、000Hzで集められた16ビット解像度データのような)高帯域、高解像度音声データ・セットが図1のブロック180への入力となる。

【0029】入力音声データ・セット180は、最初に、図1のブロック182の中で[2]で記述の再標本化プログラムを使用して、8、000Hzに再標本化される。このデータは、図1の機能ブロック182で、参考文献[8]で記述のMAXFLATルーチンの修正版を使用して設計された符復号器帯域沪波器に供給される。この沪波器は、図2、3および4の中で図示される。この沪波器の通過帯域特性は、現代の米国における電話機器の中で使われる符号化/復号化沪波器に近似するよう設計される。通過帯域、3dbポイントおよび移行(TRANSITION)帯域幅の設定は、本発明の有効性にとって重要である。ローカル電話回線上の音声に対する良好な認識を行う認識訓練のための符復号沪波器を設計するのは可能であるが、遠隔地の電話については難しい。そのような問題を避けるために、上記の特性は、たとえば、低位の3dbポイントに対しては300Hz、上位の3dbポイントに対しては3、600Hzに設定すべきである。移行帯域幅は、それぞれ、400Hzおよび800Hzでなければならない。通過帯域は500Hzから3、200Hzになる。実際の符復号器沪波器の幅に近似するために、通過帯域リプルは、全通過帯域にわたり、1単位から0.1パーセント以上の偏差であってはならない。

【0030】スフィンクス音声認識エンジンおよびタンゴラを始めとするその他の音声認識エンジンが線形沪波器によって提示されるスペクトルのひずみを感知できる点に、注意することは重要である。スペクトルのひずみは、主要な音声認識特性(例えばケプストラム)が周波数スペクトルから引き出されるので、その通過帯域の中の平坦な周波数応答を持たない。複雑な認識作業については、いくぶん平坦な通過帯域応答からのマイナーな偏差が、本願発明者の研究室において観察され、結果として、絶対認識誤り率が数パーセント劣化した。したがって、最大平坦設計アルゴリズムは、必要である。「スペクトルの傾き」へのスフィンクス音声認識エンジンの感度が、参考文献[9]の中で指摘された。したがって、MAXFLATまたは比較的低レベル通過帯域のリップル設計は、必要とされる。

【0031】4、100Hzから8、000Hzへの再標本化率変換は、参考文献[8]の中で提供されたMAXFLATには過度な要求であり、それは、帯域通過特性が符復号器帯域沪波器に必要なとき、低通過帯域フィルターの設計のためにのみ役立てられる。このルーチンに対するデザイン特性は、0.5へマップするナイキスト周波数と1.0へマップする標本化周波数によって、正規化された周波数の3dbポイントおよび移行帯域幅を表わす2個のパラメータ、ベータおよびガンマによって与えられる。Kaiserの参考文献[8]によって、ガンマは「0.005よりあまり小さくない」値に制限されなければならないことが示唆されている。これより低い値では、使われる計算精度浮動小数点数を増やすためにルーチンの修正が要求であり、そのような沪波器の条件数は、およそガンマの2乗に反比例するので、フィルター係数バッファを200から4096に拡張する必要がある。このため、44、100Hzから8、000Hzへの変換に必要となる0.05の約1/10分の1または0.0

０５のガンマ値をもつ予波器とした。２個の低域予波器設計、低域から高域通過帯域波変換、および、低域と高域通過波の満状組合せが、必要な帯域通過特性を実現するために要求された。

【００３２】上記フィルター設計の実現によって、４４、１００Ｈｚデータは、参照文献[2]で記述される再標本アルゴリズムを使用して、図１の機能ブロック１８２の中で８、０００Ｈｚに変換され、米国長距離電話機器のための通過帯域に非常に近い符復号器通過帯域を提供する。このデータは、下記のステップ２および３に従って処理され、１６ビットの、低雑音信号となる。

【００３３】同様の通過帯域特性および速度低減削減は、この訓練テクニックの中で使われる１６、０００Ｈｚ音声サンプルのために必要であるが、例外は、移行バンド要求がそれほど要求していない点と予波器加重が、要求された通過帯域平坦度特性を達成するにはさほど必要とされない点とである。　図２、３、４で、事前訓練再標本化操作の訓練に実行されたのと同様に、符復号器予波器のインパルス(Impulse)、マグニチュウド(Magnitude)およびログ・マグニチュウド(Log Magnitude) 応答を再びを示す。

２）ダイナミック・レンジを正規化するための振幅補正（スケーリング）

音声標本は、個別に読まれて、図１のブロック１８４で、１４ビットのダイナミック・レンジにスケーリングされる。

３）Ｍｕ－ｌａｗ 圧伸

各音声標本は、図１のブロック１８６で、参照文献[7]のような公の文献でよく知られているＭｕ－ｌａｗ圧縮を使用して１６ビットの精度から８ビットの精度に引き下げられる。８ビットへ圧縮されたデータは、ふたたびＭｕ－ｌａｗ公式に従って、１４ビットへ拡大される。

【００３４】この結果、図１ブロック１８８でシミュレートされる電話チャネル音声データ・セットになる。これは、信号強度によって増大、減少する量子化ノイズ・レベルを持ち、およそ一定のＳ／Ｎ比を維持する。特に、話し手の声が大きい場合、これは、電話音声信号の中で聞きとれる「ひび割れ」雑音を導入する。

【００３５】電話データより高域の種々の帯域幅で集められるであろう音声データ１８０のこのような処理は、電話機器での使用のため音声認識装置５０をブロック１９０で訓練するために使用される。音響訓練１９０は、図１のブロック１９２の音素モデルと量子化コード・ブック１０５を生成する。これにより、スフィンクス音声認識エンジンを使用して電話帯域幅での実際的音声認識を行うことを可能とする。

【００３６】シミュレートされた電話チャネルデータ使用の認識装置訓練

２個のコード・ブック１０５と２個の音素モデル・セット１９２が作成されるように、２つの訓練セッション、すなわち電話と高帯域に対するセッションが、実行される。高帯域、ローカルな認識あるいは、電話帯域幅などのユーザの要求に応じてコード・ブック１０５の各セットおよび各音素モデル１９２は、別々にに保管され、実行される。いずれの帯域幅でも、自己相関係数は、ケプストラム係数を引き出すために抜き出される。そのフレームにもっとも近い係数を類別するために、ケプストラム係数がベクトル量子化１０４によって実行される。このようにして、[1]で記述されるように、各音声時系列フレームは、そのフレームを表わす３バイトに減じられる。

【００３７】量子化の値のセットが、ビーム・サーチ・プロセス１０６に送り出される。ビーム・サーチ１０６は、ビタービ(Viterbi)ビーム・サーチと呼ばれる文法ガイド型「隠れたマルコフ・モデル」(Hidden Markov Model)サーチ・プロセスである。この文法ガイド型サーチは、サーチ・スペースを減らすために語対文法を使う。

【００３８】本発明のもうひとつの重要な点は、その音声認識システムがローカルであろうが遠隔地であろうが、両方の音声を処理することができることである。これは、音声のいずれのタイプもチャネル・シミュレータで使われる帯域幅に対応するように、実行時データ条件づけ・速度変換沢波器の遮断ポイントを２個の帯域幅の幅が狭い方に近い帯域幅に置くことによって、達成される。３dbポイントおよび移行帯域特性は、訓練の中で使われる電話符復号器沢波器の上位移行帯域の特性に近似しなければならない。

【００３９】ビーム・サーチ１０６は、そのベクトル量子化の中で引き出された時系列を語対文法からの語列に突き合わせ、各文脈を定義する。音声認識サーバは、ユーザ適用業務または音声認識クライアント（ブロック１１０）とコミュニケートする。本発明の構造は、単数のバックエンドとコミュニケートする複数のフロントエンド（ワークステーション）または複数のバックエンドとコミュニケートする複数のフロントエンドを持つことができる。

【００４０】本発明のシステムは、オペレーションの異なるレベルのために構成され実行される。非常に高いデータ速度をもつコミュニケーション・ネットワークについては、フロントエンドでのデータ圧縮のために、音声標本は、直接バックエンドを実行しているシステムに伝達されることができる。原ディジタル音声データストリームが、複数のユーザ用のバックエンドがあるサーバに送り出されることができる。電話システムについては、複数のチャネルが１つのバックエンドへつながるか、または、複数のユーザが、フロントエンドおよびバックエンド双方にコミュニケートする。

【００４１】本発明でのシステムは、音声認識サーバーとして配備される音声認識機能を中心に主として構成さ

11

れる。システムは、その時点の文脈として適用業務が選択する語対文法によってガイドされる。音声認識適用業務は、初期値設定プロシージャ、ステータス・コードおよびコマンド[6]のような機能をサポートする適用業務プログラム・インタフェース（ＡＰＩ）コールをもつ音声認識システムにインターフェースを持つ。音声認識適用業務は、音声認識サーバに一定のタイプの操作を要求するか、あるいは、ある特定の認識文脈をロードして、必要なとき、音声認識のための文脈を起動するよう要求する。音声認識適用業務が最初に実行されるとき、タスクは通常サーバによって事前ロードされる。適用業務の活動の必要に応じて、タスクはその後順に起動される。

【００４２】音声認識サーバ（ブロック１０８）のＡＰＩコールは、ユーザ適用業務（ブロック１１０）が音声認識システムのサービスを要請することを可能にする。ユーザ適用業務プログラム（ブロック１００）は、音声認識サーバの種々の構成要素と同じコンピューターまたは異なるコンピューターの上で実行することができる。同じコンピューター上の場合、適用業務プログラム（ブロック１１０）は、そのオペレーティングシステムでサポートされる共有メモリおよびセマフォを通して音声認識サーバとインターフェースをとることができる。異なるコンピュータ上の場合、交信はＲＳ２３２インターフェースあるいは遠隔プロシージャ呼出し（ＲＰＣ）を通して行われる。ＲＰＣは参照プログラミング文献[10]でよく知られている。

【００４３】ユーザ適用業務の典型的例には、エグゼクティブ情報システム、言葉の照会経由のデータベース・アクセス、ソフトウェア問題報告システムなどがある。

【００４４】もうひとつの例は、その利点を活用するため音声認識サーバへの呼び出しを行う電話回答音声応答装置（ＶＲＵ）である。ＲＩＳＣ ＳＹＳＴＥＭ ６０００（ＴＭ）およびＯＳ／２（ＴＭ）をもつ ＰＳ／２（ＴＭ）の上でこれらのサーバーは実行された。

【００４５】Ｄｉｒｅｃｔ Ｔａｌｋ ６０００（ＴＭ）は、同様の電話ＶＲＵシステムである。このＶＲＵシステムでは、１本の電話回線を扱うのではなく、（同時に活動中となる可能性のある２４個の会話チャネルをもつ）Ｔ１回線処理が必要となる。　音声認識サーバ構造は、Ｄｉｒｅｃｔ Ｔａｌｋ（ＴＭ）のように大量の電話適用業務の処理が必要な場合、複数のクライアントを扱うことができる。　ユーザ適用業務は多くの文脈を前もって登録することができる。レストラン案内、ハードディスク・ヘルプ・デスク、あるいは、ソフトウェア・ヘルプ・デスクは全て複数の文脈を階層的に事前に登録することができる。各適用業務では、何人かのユーザが、音声ストリームを入力することができる。各適用業務は、特有の音声ストリームのために特有の文脈の下で音声認識を実行するよう音声認識サーバに指示する。

【００４６】言い換えると、同じＡＰＩを扱う複数のユ

12

ーザが、１またはおそらくはいくつかの版の音声認識サーバを用いるタスクすべてを登録するであろう。システムは、要請された作業がすでにロードされているかを検査し、複数のユーザの音声認識タスクが余分にロードされることを回避する。

【００４７】タスク構築（ブロック１３０）は、いくつかの基本入力ソースを持つ。２０，０００語の発音をもつ基本辞書である米語辞書（ブロック１３２）は、その１つである。補足辞書（ブロック１３８）は、適用業務特有のもので、基本辞書の中で見つけられなかった語の発音を追加するためのものである。補足辞書は、典型的には、特定の適用業務が音声認識のために必要とする固有名詞、頭字語(ACRIMYM)その他から構成される。

【００４８】基本米語辞書（ブロック１３２）は、タスク構築プログラム（ブロック１３４）によって求められる語および音素を供給する。タスク構築プログラムは、また、何がそのタスクの下の音声認識サーバによって認識されることができるかを決めるためにタスクＢＮＦ辞書（ブロック１３６）から該当するタスクＢａｕｋｕｓ－Ｎａｕｒ Ｆｏｒｍ（ＢＮＦ）文法を引き出す。たとえば、地域レストラン情報を提供する適用業務の最初の文脈は、その話し手が希望するレストランのタイプ、たとえば、フランス、イタリア、中国料理などであるかもしれない。ひとたびそのタイプが決まれば、次の文脈は、その特定のカテゴリの中のレストランとなろう。タスク構築プログラムは、そのパターン合わせのために必要なすべての語を見つけるためにＢＮＦを分析し、汎用の米語辞書(ブロック１３２)から音素表示を引き出す。必然的に、あらゆる特定適用業務は、そのシステムに加えられなければならない それ自身の副語彙を持ち、それらは、補足辞書に保存される。たとえば、レストラン・ヘルプ・デスクの中で、「イタリアン」、「フレンチ」、「スパニッシュ」などの言葉は、汎用米語辞書で見つけられるが、レストラン名、とくに外国語で、たとえば、「Cherchez LesFemmes」、「Chateau Voulez」や、アメリカのレストランで普通でない名、たとえば、Ｊ．Ｊ．Ｍuldoon、は、普通の辞書になく、タスク補足辞書（ブロック１３８）に加えなければならない。これらの補足辞書（ブロック１３８）は、また、基本汎用米語にあるが発音をローカルなものにするためにローカルな語彙を含めることができる。。

【００４９】タスク構築プログラム（ブロック１３４）は、入力ＢＮＦ文法を分析して、その文法の中の各語のリストと次に続くことができるすべての語のサブリストを生成する。したがって、その文法の中の各語が、後に続く適切な語のリストおよび各語の音素表示へのポインタを持つ。音素モデル１９２は、種々のＶＱ値を観察するである。このマルコフ・モデルは、ＶＱ値（ブロック１０４）のための、一群の離散的確率分布であり、「隠れたマルコフ」状態機械が音素の範囲内の特定の状態に

あるとすると、VQ値のオカレンスの確率を与える。「隠れたマルコフ・モデル」は文献[11]に適切に記述されている。

【0050】ビーム・サーチ（ブロック106）は、訓練プロセスの間に生成される文脈感知のトリフォン(triphones)の大きいテーブルから連結HMM音素モデル192でできている語モデルを使用する。この語モデルが、VQ値の観察された順序を最もよく説明する語順序の最適推定を行うために使われる。ビーム・サーチ（ブロック106）は、そのサーチの中で使われる語をつくるための音素モデル192を選択するために、語文法を使う。

【0051】ユーザ適用業務は、音声認識サーバを制御する。例えば、[12]で記述されるIBMプログラム・プロダクト Direct Talk/2(TM)は、電話に応答しレストラン案内機能を実行するひとつのユーザ適用業務となり得る。レストラン案内適用業務は、Direct Talk/2(TM)を使用し、この適用業務が16の文脈を持ち、レストラン案内ヘルプ・デスクの一部である文脈を事前ロードする要求を起こすことを音声認識サーバに知らせる。その適用業務が進行するにつれて、音声認識サーバの文脈切り替えを要請する。ユーザは、電話ヘルプを電話を通して呼び出す。レストラン案内は、音声認識サーバに最初のレベルの文脈での音声認識を実行することを要請する。認識サーバとユーザ適用業務間のAPI上で制御とデータが交換される。Direct Talk/2 (TM) システムの複数の事例が同じ音声認識サーバを使用する可能性がある。

【0052】音声認識サーバは、無声間隔(ユーザが調整可能で、ほとんど一般に0.6秒）が来るまで音声データを捕捉する。無声間隔が観察されると、認識は終了し、話し手の話しが終わったと仮定される。

【0053】本発明記載の音声認識システムは、複数のハードウェア・プラットホームおよび複数のソフトウェア機械構成の上に、複数の実施を可能にするよう基本設計がなされる。たとえば、1つの可能な構造は、図5のように、ローカル・エリア・ネットワーク160を通して接続されているワークステーションの物理的実施の上への上記論理的構造50の物理マッピングを提供する。この構造の中の各ワークステーション150、150'、150"は、複数の独立ユーザ適用業務を実行することができ、各々は、スレーブ・プロセッサーとしての音声認識サーバ50のマスターとなる。PBX170は、外部の電話回線に接続していて、電話帯域幅データストリームを図1の音声認識サーバ50のアナログ・デジタル変換100に渡す。認識された音声を表わしているテキストが、ワークステーション150、150'、150"のユーザ適用業務に音声認識サーバから返される。

【0054】訓練プロセス

ビーム・サーチ・プロセスでの語モデルとテキストとのパターン合わせに使われる音素HMM192のパラメータを推定するために、訓練プロシージャは、既知の音声およびテキスト原稿という大規模辞書を使用する。

【0055】最初に、その原稿が、訓練セットの語の発音を表わす音素を汎用米語辞書から検索するために使われる。

【0056】次に、音素HMM192のパラメータが、共調音(coarticulation)効果の効果的な推定を行うために、先行および後続音素文脈（トリフォン-triphones-と呼ばれる）の中で推定される。使われる推定プロシージャは、[1]で記述のBaum-Welch 順方向／逆方向繰返しアルゴリズムである。訓練されたトリフォンHMMが訓練セットの中で観察されたVQ値時系列を生成したであろう確率を最大にするために、HMMのパラメータが、繰り返し調節される。

【0057】あらゆる「隠れたマルコフ」音素モデルには多くのパラメータがあり、各「隠れた」状態機械中に7つの状態および12のトランジション・アーク(TRANSITION ARC)が存在する。各トランジション・アークに関連して、3つのコード・ブックの各々の確率分布に、関連する256の離散エレメントがある。訓練プロシージャから生じるトリフォンHMMパラメータは、連続音声の中に存在する共調音効果を表わすのに必要なトリフォン数を減らすために一定の幅の値の範囲に集められる。

【0058】訓練は、ローカルな電話交換を通して集められる低レベル帯域幅音声およびマイクからの高帯域音声の組合せによって実行される。高帯域音声は、本発明に従って、本願書で記述の電話チャネル・シミュレータ185によって処理される。3個のコード・ブックすべては、この段階でコンパイルされる。[1]で記述のように、コード・ブックが、ケプストラム、微分のケプストラム、およびべきおよび微分のべきを含む。

【0059】3つのコード・ブックの各々は量子化コード・ブック105に保存され、実行時ベクトル量子化プロセスで使われる。ここで、電話ネットワークの効果が、データの事前処理によってシミュレートされ、公衆電話ネットワークが調整するのと同じ方法で特性コード・ブックの統計的属性が調整される。このプロシージャをとることによって、米国大陸の様々なロケーションからの呼び出しをもつ実際の電話の音声認識の正確度が大幅に増加した。

【0060】図6は、たとえばPBX170経由で電話から得られた音声に応答する音声認識装置50を訓練するための電話チャネル・シミュレーション・プロセス200を記述する流れ図である。図6の流れ図は、図5のデータ処理装置50の上で実行されることができるコンピューター・プログラム方法を表わす。

【0061】プロセス200は、電話帯域幅より帯域幅が広い音声認識訓練プロセッサ50に音声データ・

15

セットを入力するステップ２０２で始まる。例となる高帯域音声データ・セットは、参照文献[14]から[19]で記述されている。このステップは、図１のデータ入力ブロック１８０に対応する。

【００６２】図６のステップ２０４で、音声データ・セットは、電話帯域幅を持つ間引かれた音声データ・セットを得るために間引かれる。これは、図１の機能ブロック１８２に対応する。間引かれた音声データ・セットは、入力音声データ・セットの高い方の帯域幅より低い帯域幅を持つであろう。間引き(decimation)プロセスは、参照[2]で記述されている。

【００６３】次に、図６のステップ２０６で、帯域通過デジタル沪波器を間引かれた音声データ・セットに適用し、電話機器の伝送特性に特徴づける。これは、図１の機能ブロック１８２に対応する。これは、沪波された音声データ・セットを得るために行われる。帯域通過デジタル沪波器は、最大平坦設計アルゴリズムを持たなければならない。

【００６４】次に、図６の中のステップ２０８で、その最大ダイナミック・レンジが非圧伸電話音声の最大レンジと一致するように、沪波された音声データ・セットの振幅が、再補正される。これは、図１の機能ブロック１８４に対応する。これは、振幅再補正音声データ・セットを得るために行われる。このステップの結果、その最大ダイナミック・レンジは非圧伸Ｍｕ－ｌａｗ電話音声の最大ダイナミック・レンジと一致し得る。代わりに、その最大ダイナミック・レンジは非圧伸Ａ－ｌａｗ電話音声の最大ダイナミック・レンジと一致することもできる。

【００６５】次に、図６ステップ２１０で、上記補正音声データ・セットを、電話中の音声信号の圧伸非圧伸の順序を表わしている量子化ノイズをもって修正する。これは、図１の機能ブロック１８６に対応する。これは、修正された音声データ・セットを得るために行われる。修正ステップは、Ｍｕ－ｌａｗノイズとしての量子化ノイズを持つことができる。代わりに、修正ステップは、Ａ－ｌａｗノイズとしての量子化ノイズを持つことができる。

【００６６】次に、図６のステップ２１２では、統計的パターン・マッチング・データ装置を訓練するために、音声認識プロセッサ５０へ修正された音声データ・セットを入力する。これは、図１の出力データ・ブロック１８８に対応する。シミュレートされた電話チャネル音声１８５が、電話音声特有性を持つ電話コード・ブック１０５の特性を持つ音楽モデル１９２を生成するために、音響的訓練プロセス１９０によって使われる。

【００６７】次に、図６のステップ２１４で、たとえば、図５のＰＢＸ１７０からの信号のような、電話からの音声信号に対し、音声認識プロセッサ５０を使って、音声認識が実行される。

16

【００６８】電話チャネル・シミュレータ（ブロック１８５）を使用する高帯域音声の変換は、連続の音声認識装置に限られてなく、たとえば、IBM Tangora Dictation SystemおよびDragon Systems、ニュートン・マサチューセッツ、Dragon 30k DictateおよびKurzweil Applied Intelligence、Voice Report、Waltham、マサチューセッツおよび[20]で記述されるその他のシステム等のような様々な音声認識プロセッサに適用されるということに留意する必要がある。

【００６９】上記本発明の説明において引用した参照文献は、以下の通りである。

【００７０】[1] "Large Vocabulary Speaker and Dependent Continuous Speech Recognition: The Sphinx System"; Kai-Fu Lee; Cargnie Mellon University, Department of Electrical and Computer Engineering; April 1988; CMU-CS-88-148

[2] "A General Program to Perform Sampling Rate Conversion of Data by Rational Ratios"; from "Programs for Digital Signal Processing", Ed.: Digital Signal Processing Committee of the IEEE Acoustics, Speech, and Signal Processing Society; IEEE Press, 1979; Section 8.2, pp8.2-1 to 8.2-7 by R.E. Crochiere

[3] "Theory and Application of Digital Signal Processing" L.R. Rabiner, B. Gold; Prentice Hall, 1975, pp 91

[4] "Digital Processing of Speech Signals"; Prentice Hall Signal Processing Series; 1978, pp 401-402, 411-413

[5] "An Algorithm for Vector Quantizer Design"; Y. Linde, A. Buzo, R.Gray, IEEE Transactions on Communications, Vol. com-28, no. 1, January 1980

[6] "IBM Continuous Speech Recognition System Programmers Guide"; B. Booth; 1992; currently unpublished, available on request.

[7] "Digital Telephony and Network Integration"; B. Keiser, E. Strange;Van Nostrand Reinhold Company Inc. 1985.; pp. 26-31

[8] "Design Subroutine (MAXFLAT) for Symmetric FIR Low Pass Digital Filters with Maximally-Flat Pass and Stop Bands" from "Programs for DigitalSignal Processing", Ed.: Digital Signal Processing Committee of the IEEE Acoustics, Speech, and Signal Processing Society; IEEE Press, 1979; Section 5.3, pp 5.3-1 to 5.3-6 by J. Kaiser

[9] "Acoustical and Environmental Robustiness in Automatic Speech Recognition" A. Acero; Cargnie Mellon University, Department of Electrical andComputer Engineering; April 1990; CMU-CS-88-148

[10] "AIX Distributed Environments: NFS, NCS, RPC,

DS Migration, LAN Maintenance and Everything"; IBM International Technical Support Centers,Publication GG24-3489, May 8, 1990
[11] "A Tutorial on Hidden Markov Models and Selected Applications inSpeech Recognition"; L. Rabiner; Readings in Speech Recognition; Ed.: A.Waibel, K. Lee; Morgan Kaufmann; 1990; pp 267-296
[12] "IBM CallPath DirectTalk/2 General Information and Planning Manual"; International Business Machines publication no. GB35-4403-0; 1991
[13] "A Maximum Likelihood Approach to Continuous Speech Recognition";L. R. Bahl, F. Jelinek, R. Mercer; Readings in Speech Recognition; Ed.:A. Waibel, K. Lee; Morgan Kaufmann; 1990; pp 308-319
[14] "Speech Corpora Produced on CD-ROM Media by The National Institute of Standards and Technology (NIST)", April, 1991
[15] "DARPA Resource Management Continuous Speech Database(RMI) Speaker Dependent Training Data", September 1989 NIST Speech Discs 2-1.1, 2-2.1 (2 Discs) NTIS Order No. PB89-226666
[16] "DARPA Resource Management Continuous Speech Database(RMI) Speaker-Independent Training Data", November 1989 NIST Speech Disc 2-3.1 (1 Disc) NTIS Order No.PB90-500539
[17] "DARPA Extended Resource Management Continuous Speech Speaker-Dependent Corpus (RM2)", Sepetember 1990 NIST Speech Discs 3-1.2, 3-2.2 NTIS Order No. PB90-501776
[18] "DARPA Acoustic-Phonetic Continuous Speech Corpus (TIMIT)", October 1990 NIST Speech Disc 1-1.1 NTIS Order No. PB91-0505065
[19] "Studio Quality Speaker-Indpendent Connected-Digit Corpus(TIDIGITS)", NIST Speech Discs 4-1.1, 4-2.1, 4-3.1 NTIS Order No. PB91-505592
[20] "The Spoken Word", Kai-Fu Lee, et al., Byte Magazine, July 1990, Vol- 15, No. 7; pp. 225-232

【００７１】

【発明の効果】電話回線から入力される不特定の話し手の音声を音声認識するシステムを構築することによって、たとえば、電話による顧客問い合わせ自動応答システムやレストラン電話案内など、従来技法では実現できなかった新たなコンピュータ適用業務を開発することができる。

【図面の簡単な説明】

【図１】電話チャネル・シミュレータ発明を含む、連続音声認識システムの論理的構造を図示する。

【図２】電話の符復号器沪波器インパルス応答を特徴づけるグラフである。

【図３】振幅特性頻文規格化ラジアン周波数を図示するグラフである。

【図４】対数振幅特性頻文規格化ラジアン周波数を図示するグラフである。

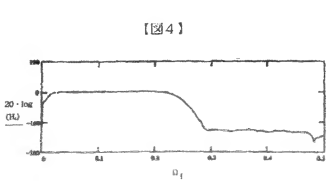
【図５】電話顧客業務通話センタにおける音声認識サーバのネットワークのブロック図である。

【図６】電話から得られる音声に応答するために音声認識装置を訓練するためのプロセスのステップ流れ図である。

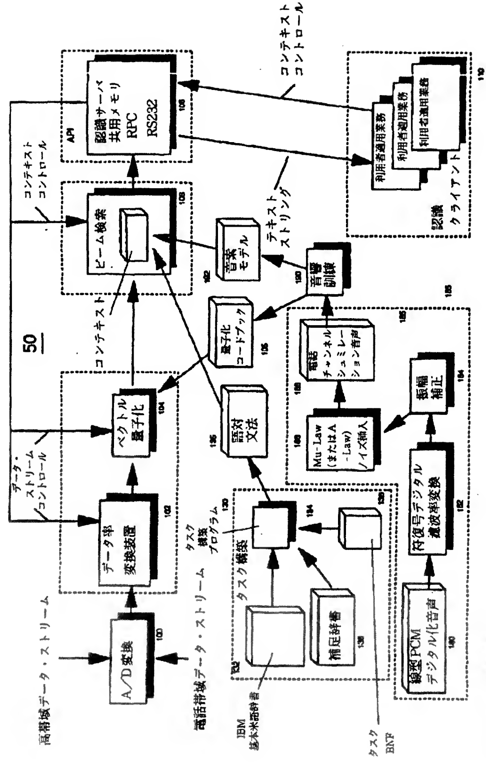
【符号の説明】

１００　アナログ・デジタル変換
１０４　ベクトル量子化
１０５　ベクトル量子化コードブック
１９２　音素モデル
１３５　語対文法
１３２　米語辞書
１３８　補助辞書
１８６　Ｍｕ－ｌａｗノイズ
１８６　Ａ－ｌａｗノイズ
１８８　電話チャネル・シミュレータ
１８２　符復号デジタル沪波・速度変換
１８４　振幅補正（スケーリング）
１３４　タスク構築プログラム
１０６　ビーム・サーチ
１０８　ＡＰＩ（適用業務プログラム・インターフェース）

【図４】

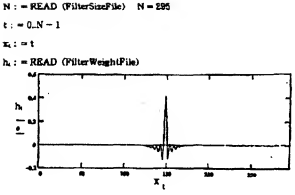


【図1】

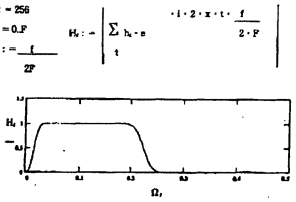


ICSRS システム論理構造

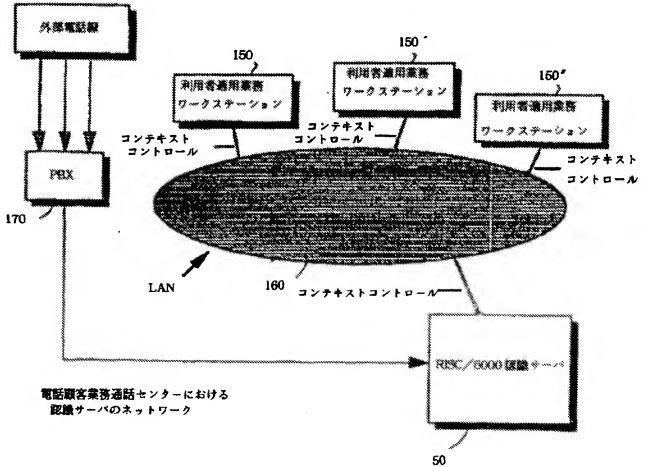
【図2】


N : = READ (FilterSizeFile)　N = 255
t : = 0..N − 1
x_t : = t
h_t : = READ (FilterWeightFile)
h_t
x_t

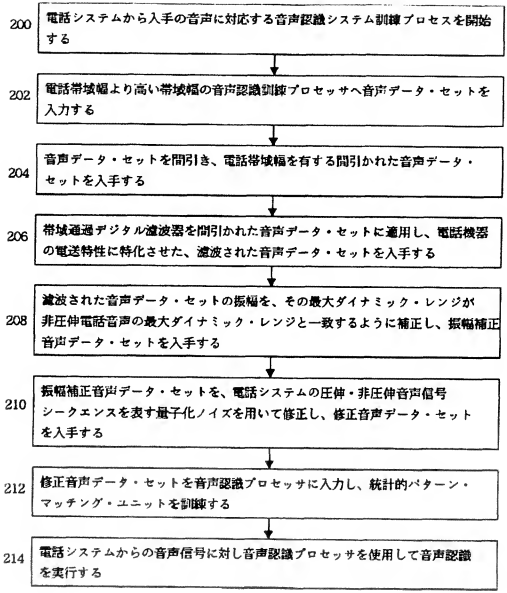
【図3】


F : = 256
f : = 0..F
Ω_f : = f / 2F
H_f : = |Σ_t h_t · e^(−i·2·π·t·f/2·F)|
H_f
Ω_f

【図5】


外部電話線
PBX
170
150
利用者適用業務
ワークステーション
150'
利用者適用業務
ワークステーション
150"
利用者適用業務
ワークステーション
コンテキスト
コントロール
コンテキスト
コントロール
コンテキスト
コントロール
LAN
160
コンテキストコントロール
RISC／6000 認識サーバ
50
電話顧客業務通話センターにおける
認識サーバのネットワーク

【図6】



フロントページの続き

(72)発明者　ノーマン　エフ　ブリックマン
アメリカ合衆国メリーランド州ボトマック
ミルバン　ドライブ　11709番地